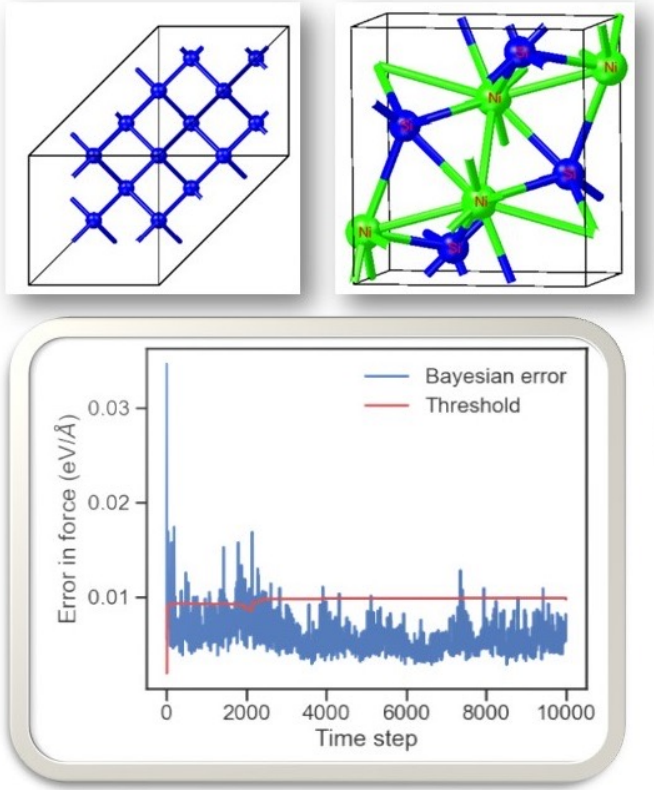
materials design

On-the-fly Machine Learning Forcefields with *MedeA VASP*



Shubham Pandey and Xiaoli Liu

Materials Design, Inc.

20th October 2022

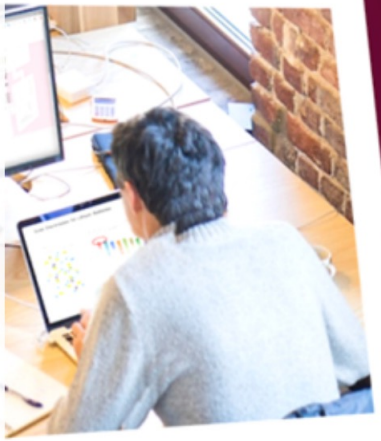
Materials Design UGM

UGM 2022

The Materials Design annual user event will be online for 2022.

Plenary Speakers include:

Prof. Jeffrey Grossman
Prof. Georg Kresse
Dr. Carla Verdi
Prof. Jörg Behler
Dr. Jozef Bicerano



https://ugm.materialsdesign.com/

Materials Design UGM Training Series

- Share the plenary sessions with your colleagues!
  - Registration details

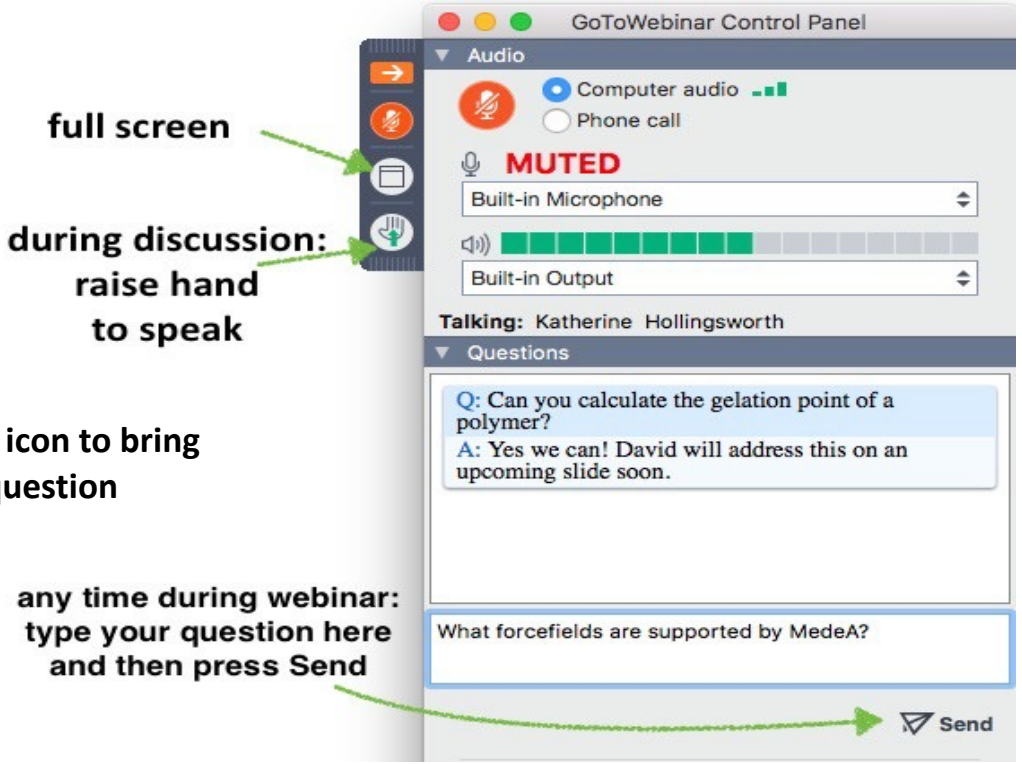
    https://www.ugm.materialsdesign.com
- We will be recording this session
  - Upcoming sessions are posted on the UGM site
  - Watch any of our earlier webinars anytime www.materialsdesign.com/webinars
- Brief survey
  - Take a 2 minutes brief survey at the end of the webinar
- Audio issues
  - Log out and log back in again
  - Check your audio output
  - Google Chrome (most recent 2 versions) Mozilla Firefox (most recent 2 versions) Apple Safari (most recent 2 versions) Microsoft Edge (most recent 2 versions)

materials design

Training & Support Team

Xiaoli Liu	Shubham Pandey	René Windiks	Thomas Nilson	Garrett Tow	David Reith
Presenter	*Presenter*				*Moderator*

Please Ask Questions!



Use the raise hand icon to bring attention to your question

any time during webinar: type your question here and then press Send

materials design

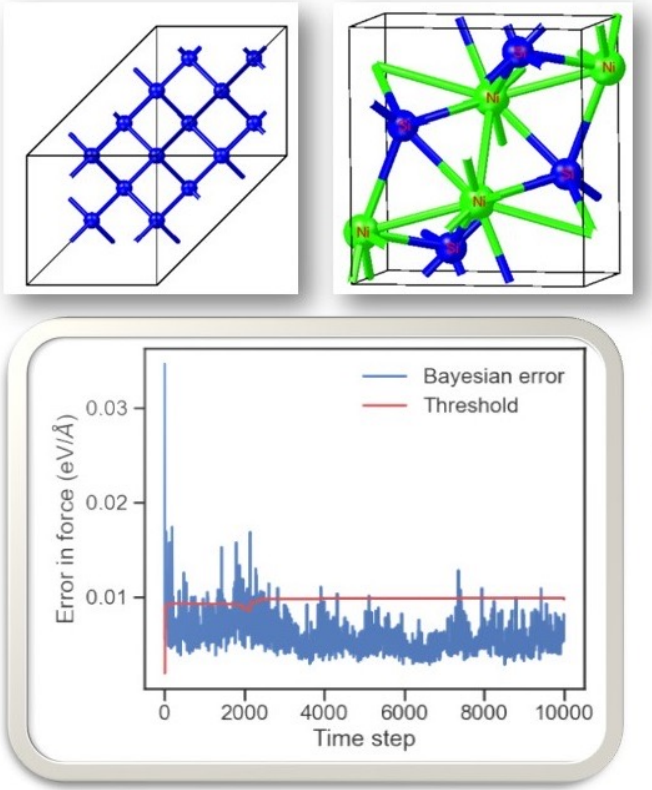
materials design

On-the-fly Machine Learning Forcefields with *MedeA VASP*

Shubham Pandey and Xiaoli Liu

Materials Design, Inc.

20th October 2022



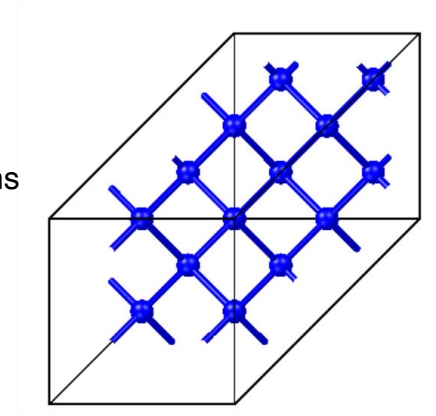
Outline

- Motivation
- MLFF basic principles
- Training MLFFs for Si, NiSi, and Si/NiSi interface
  - MLFF quality with Bayesian error and RMSE analysis
- Applying MLFF to predict volume-energy curve, thermal expansion coefficient
- Long MD simulation of Si/NiSi interface with MLFF
- Guidelines for MLFF training

Ab-initio Molecular Dynamics

- Ab-initio MD traditionally used to accurately determine dynamic properties of materials
- Interactions of atoms and electrons are calculated fully quantum-mechanically
- Limited to small simulation times (few ps) and small cells (few 100 atoms)
- One of the workarounds: machine-learned forcefields to dramatically accelerate these calculations

Pure Si, 16 atoms



nPT, 2fs tstep, 400K, 12 CPU cores

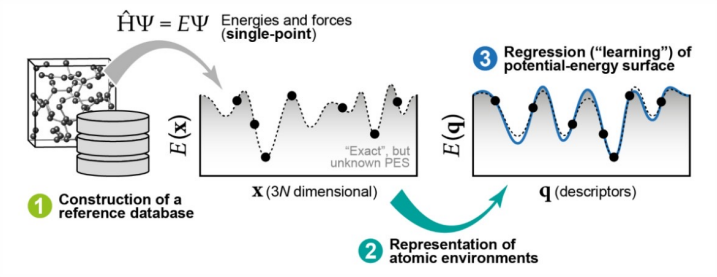
- 3ps VASP-MD: 32 min
- 20ps MLFF-MD: 1.6 min

Key steps of machine-learning

Step-1: training data generation (“on-the-fly”)

Step-2: feature selection (representation of local environment)

Step-3: ML model selection (such as neural network or a regression method)



Deringer et al., *Adv. Mater.* 31, 1902765 (2019)

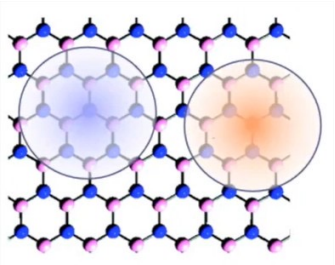
Database → Features → Regression

materials design

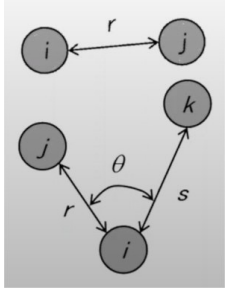
Features and regression model

Database → Feature → Regression

- Combination of pair-correlation and angular functions
  - $\rho^1(r) = g(r)$
  - $\rho^2(r, s, \theta)$
- Likelihood of finding of one particle at distance *r*, and another at *s* at an angle *θ*
- Both are translationally and rotationally invariant
- Total energy is then a function of local environment dependent local energies (just an approximation – DFT does not break up the energies into local energies!)
- Kernel regression to approximate the total energy function
- $E = \sum_i^{N_a} \sum_{i_B} K(x_i, x_{i_B}) w_{i_B}$
- “reference atoms” are chosen from the entire dataset and fitting weights are determined for those reference atoms to obtain an approximation of the total energy



Classical density distribution of atoms around a central atom



Pair-correlation and angular functions

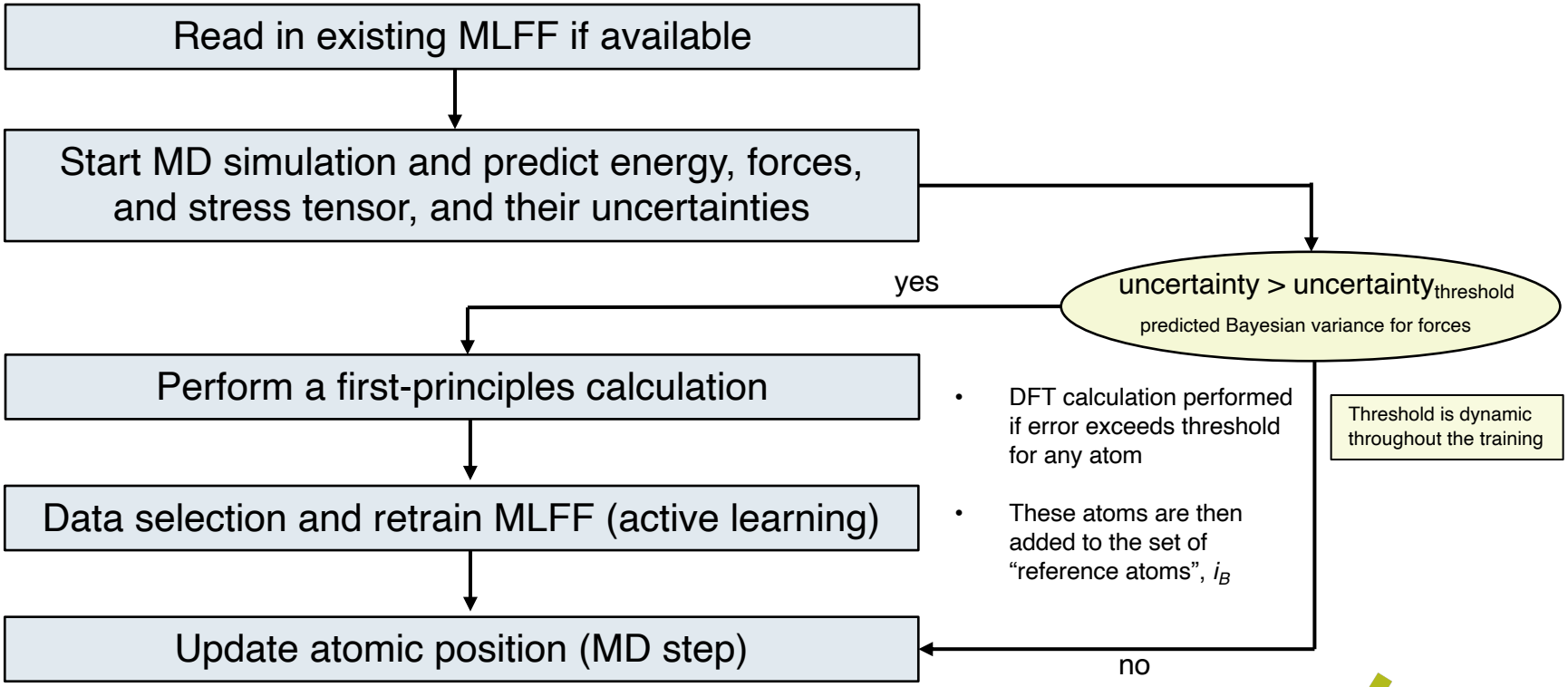
Jinnouchi et al., *J. Chem. Phys.* 152, 234102 (2020)

How do we choose training set structures and those reference atoms?

materials design

On-the-fly machine-learning

Database → Descriptors → Regression



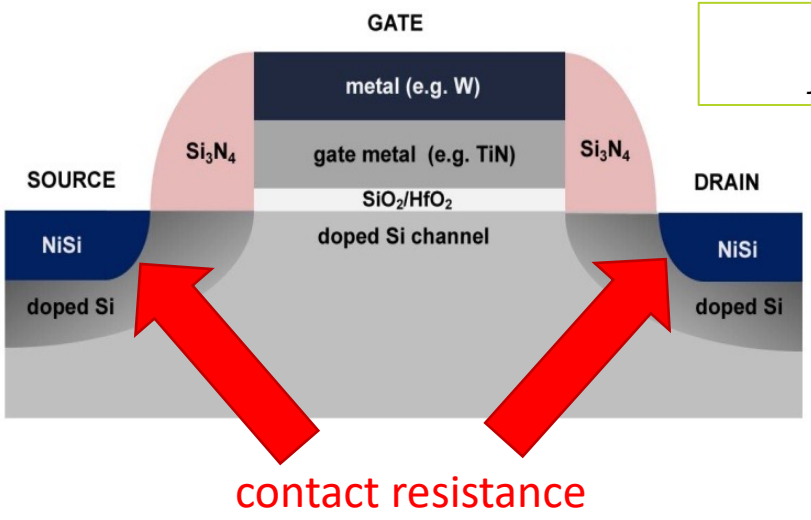
Jinnouchi et al., *Phys. Rev. B* 100, 014105 (2019)

materials design

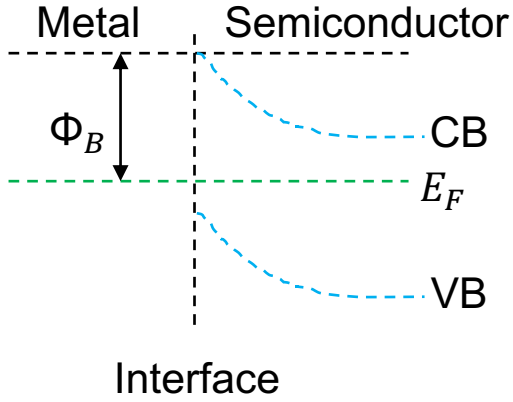
MLFF for Si, NiSi, and Si/NiSi Interface: motivation

NiSi/Si interface in CMOS device

NiSi is commonly used in complementary metal oxide semiconductor (CMOS) devices because its workfunction can be adjusted. The contact resistance at the source and drain remains a critical bottleneck, which is controlled by the Schottky barrier (SB) at the interface between the metallic NiSi and semiconducting Si regions.

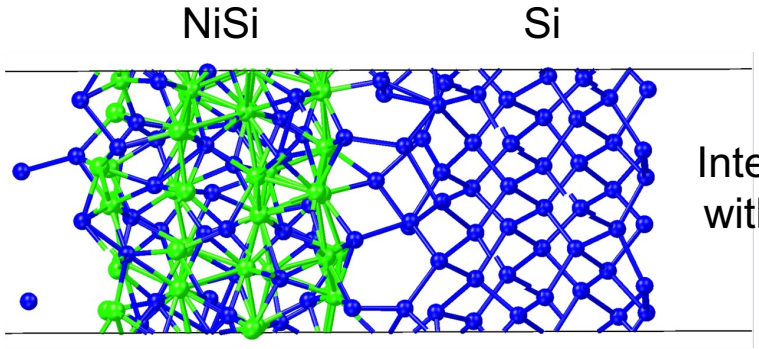


$$J \propto e^{\frac{kT}{SBH}}.(e^{\frac{qV}{kT}} - 1)$$

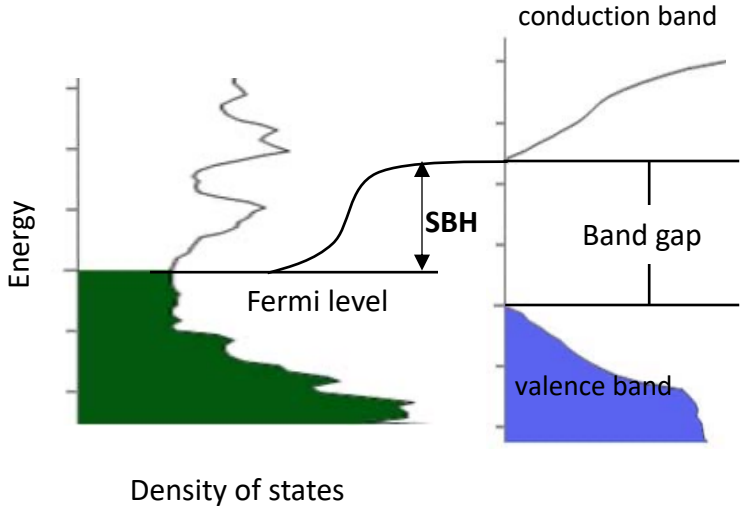


materials design

Possible application of the NiSi/Si interface structure



Interface model with MLFF-MD



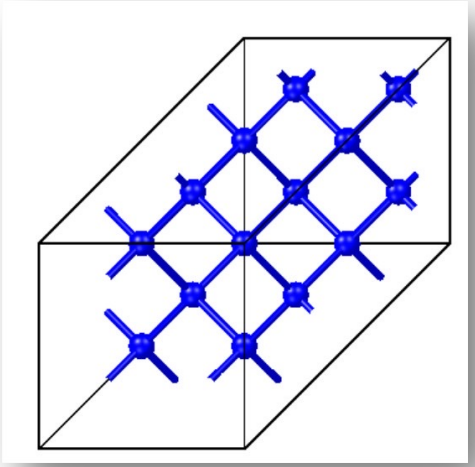
DOS alignment to predict SBH

What can MLFF do?

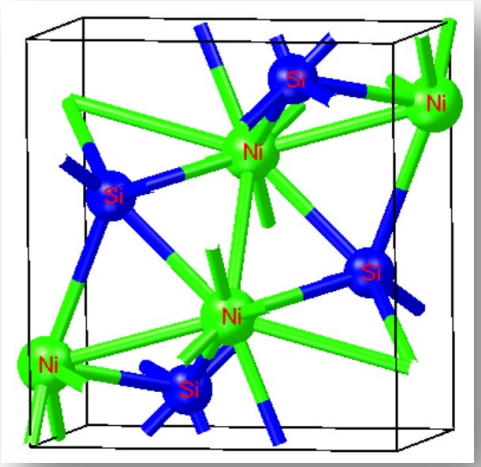
- Accelerate VASP-MD calculation
- Calculate for a larger system
- Longer simulation time (e.g., 20 ps)
- …

materials design

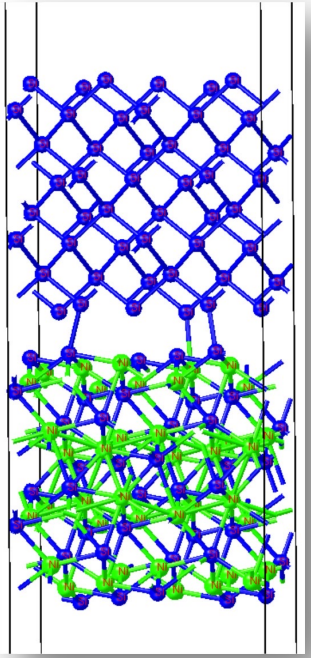
Material systems of interest: Si, NiSi, Si/NiSi interface



MLFF-1: Si



MLFF-2: Si & NiSi, Pnma, s.g. #62 (bulk)

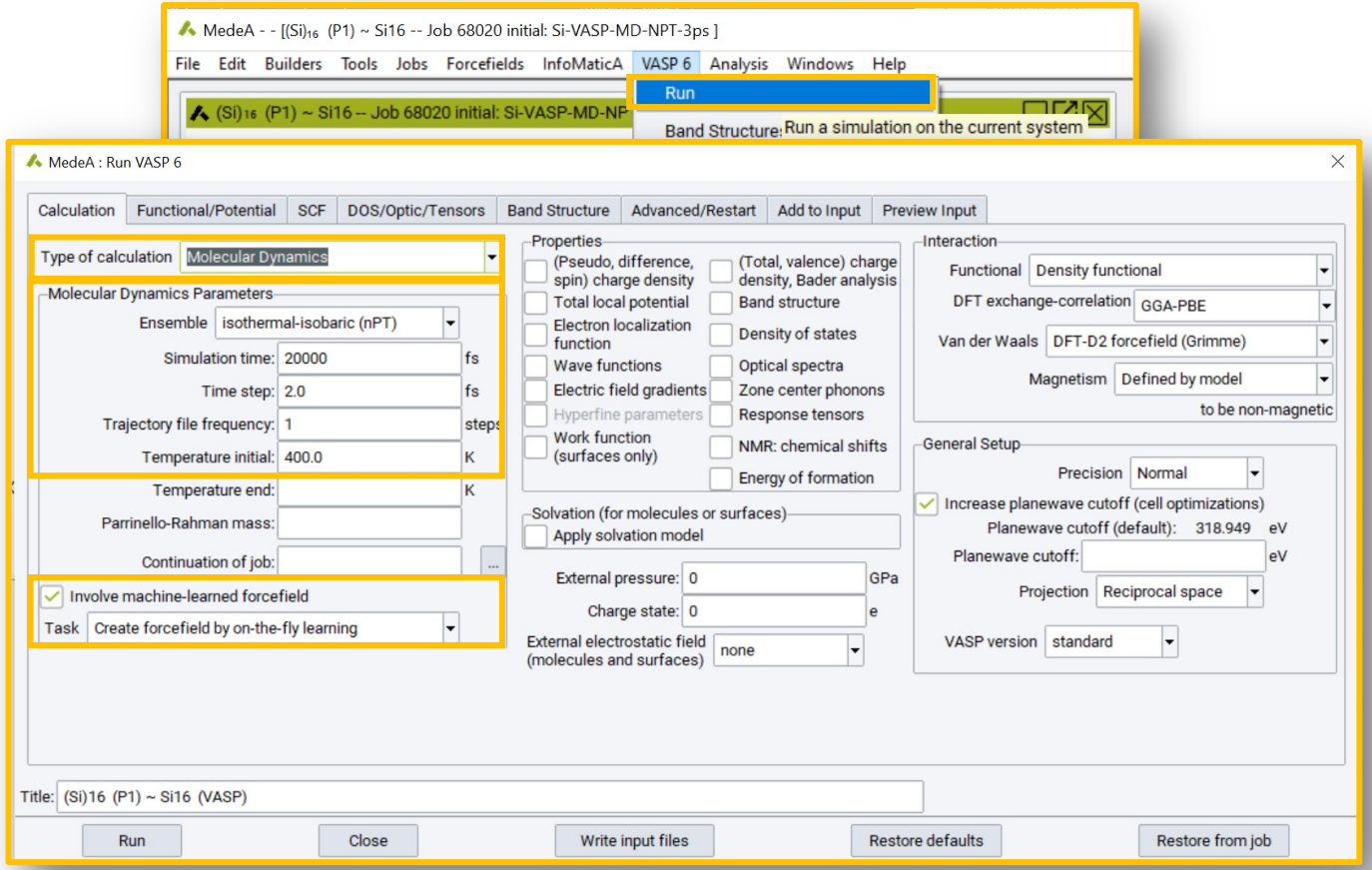


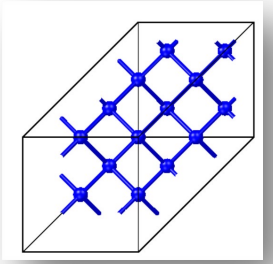
MLFF-3: Si & NiSi (bulk) & Si/NiSi(100) interface

- Step-1: train an MLFF on pure Si only
- Step-2: train an MLFF on NiSi starting from MLFF created in step-1
- Step-3: finally, train an MLFF on Si/NiSi interface starting from MLFF created in step-2

MLFF for pure Si

Setting-up MLFF training with MedeA VASP





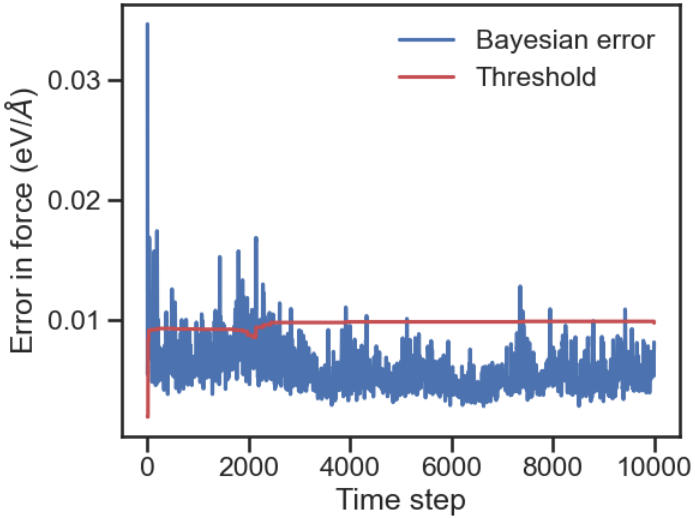
- VASP 6 GUI
- Type of calculation: Molecular Dynamics
- nPT Ensemble
- 10,000 time-steps
- Elevated temperature
- Involve machine-learned forcefield

How to judge the MLFF quality once training is completed?

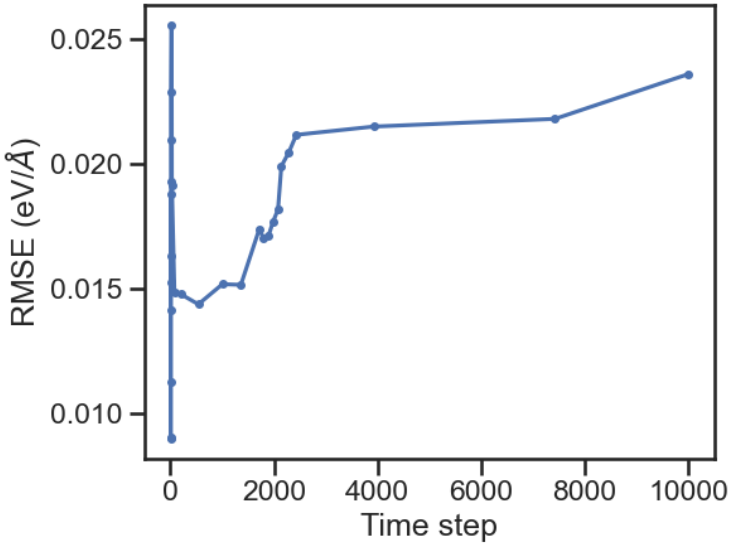
materials design

(1) Pure Si - Bayesian error and RMSE on atomic forces

Bayesian error on maximum of all forces on atoms (an estimate of *out-of-sample* error)



RMSE of forces w.r.t ab-initio data (real error on training structures up to the step where FF is updated)

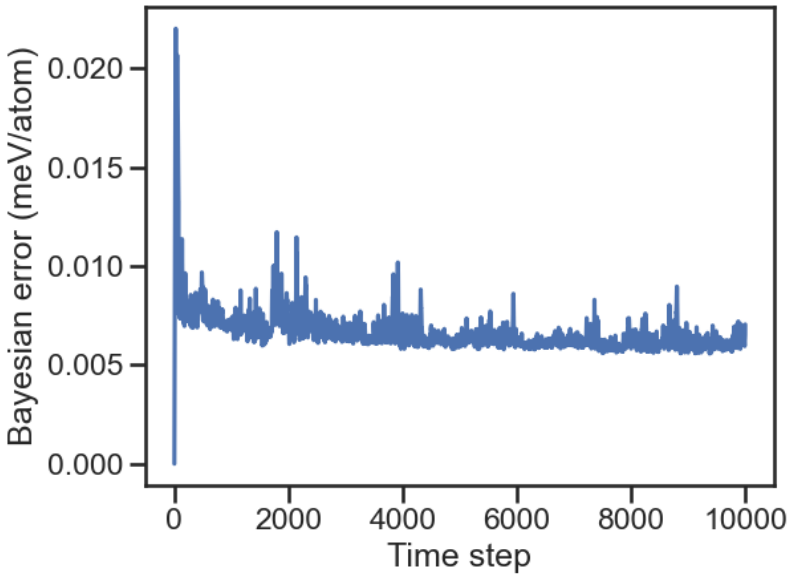


- Bayesian error catches error in the data but retaining error in the probability model
- Fluctuations in Bayesian error go down with increasing MD steps
- Sudden fluctuation at an MD step indicates a local configuration of atoms encountered by forcefield far away from what VASP has learned
- Training time: ~23 min, 16 atoms, 20 ps, 12 cores
- Good accuracy: 0.03-0.1 eV/Å for temperatures 300-1000 K

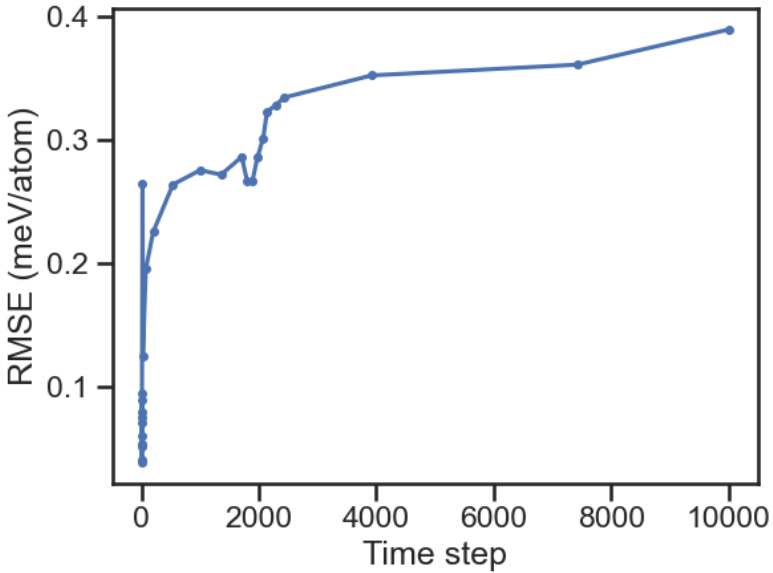
materials design

(1) Pure Si - Bayesian error and RMSE on total energies

Bayesian error of energy per atom



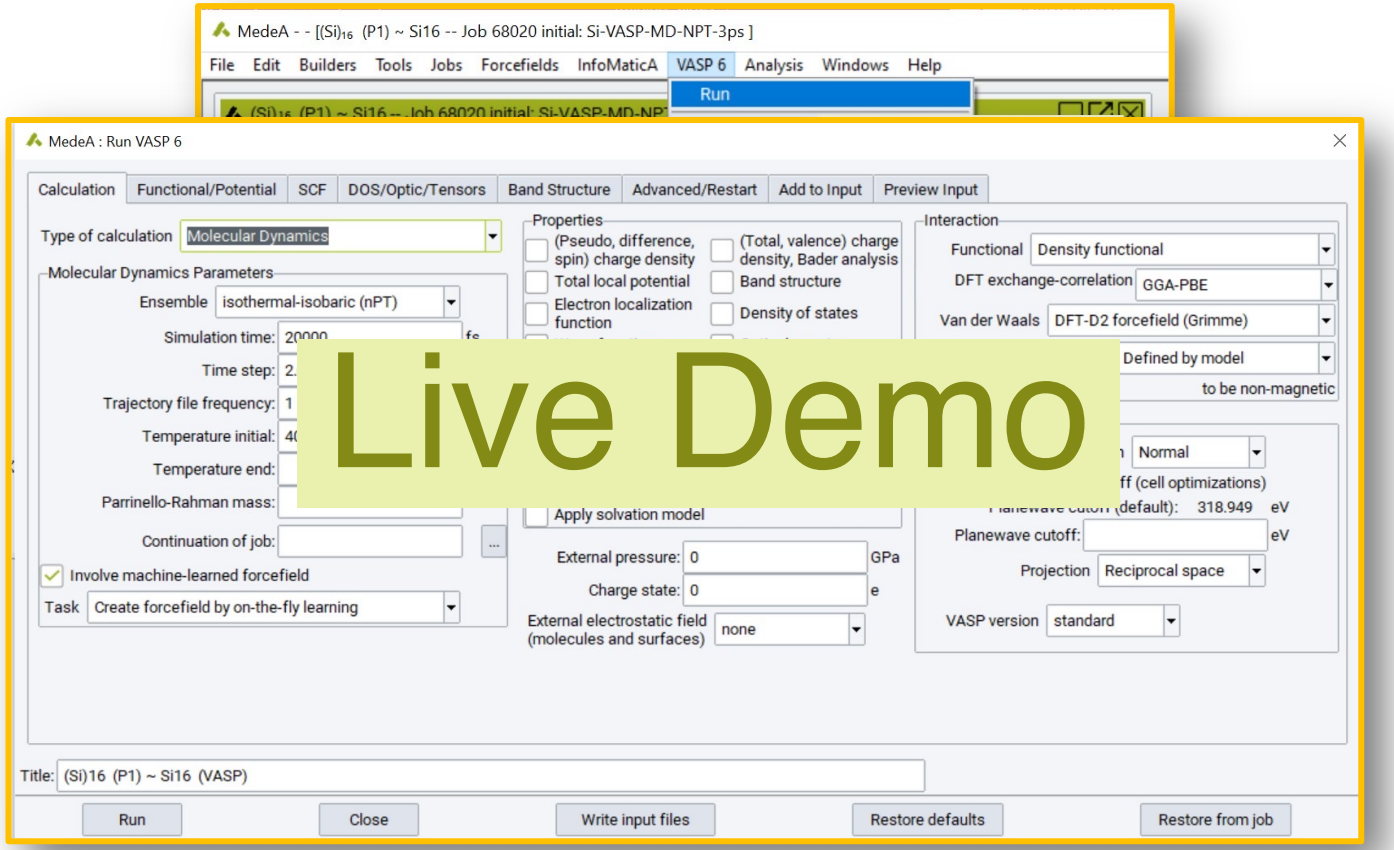
RMSE of energy per atom



Good accuracy: <5 meV/atom for temperatures 300-1000 K

How to use MLFF for property calculations?

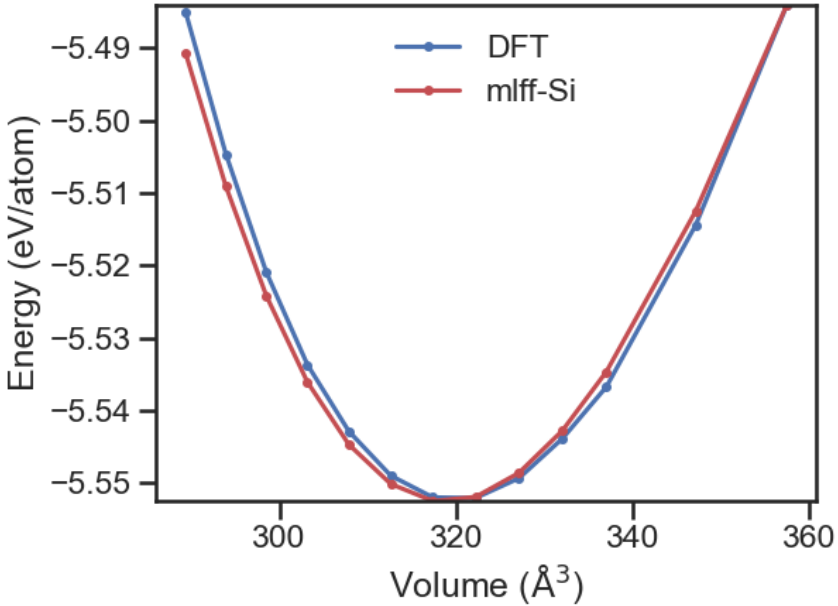
materials design

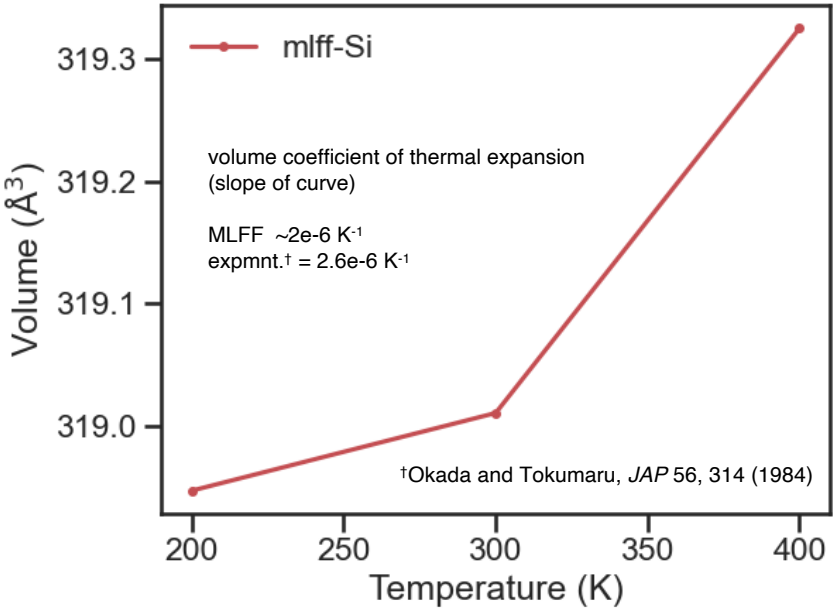


Live Demo

materials design

(1) Pure Si – volume-energy curve and thermal expansion
MLFF trained from scratch on Si

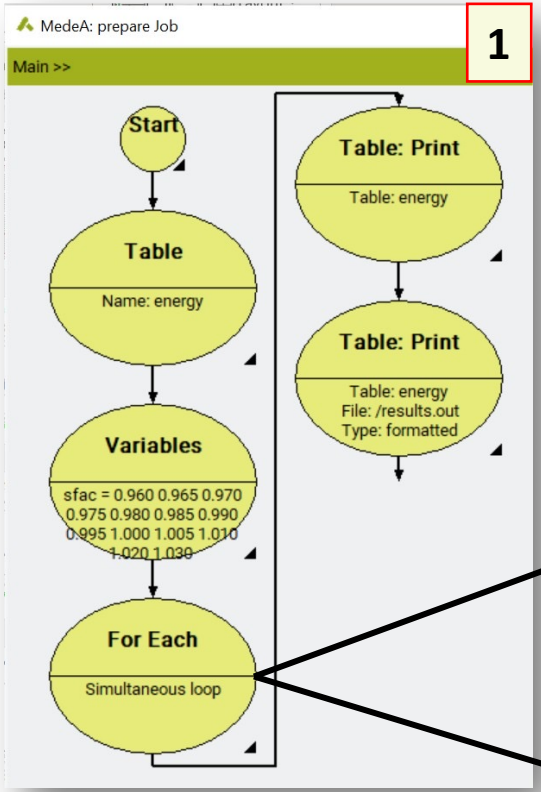


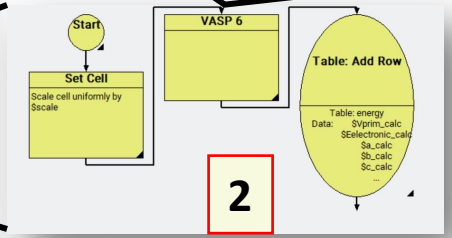


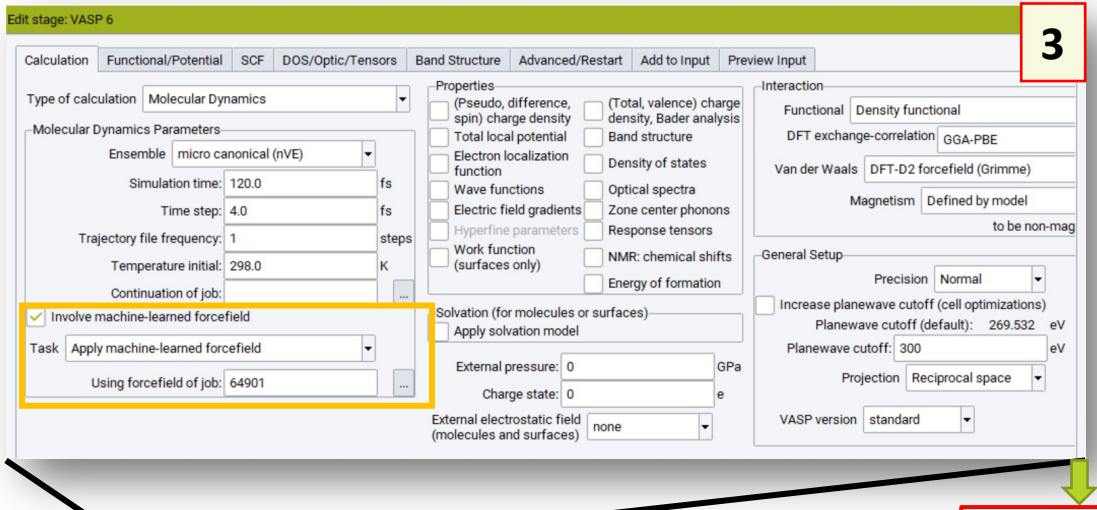
- Single-point calculations on expanded/contracted cells
- Good agreement between MLLF predicted lowest energy volume with DFT
- Fairly good agreement of MLFF predicted thermal expansion coefficient with experiments
- Compute time: ~1.6 min, 16 atoms, 20 ps, 12 cores

materials design

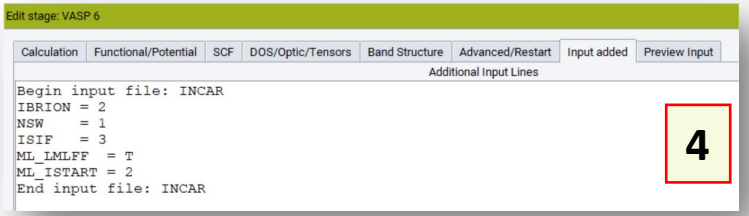
Apply trained MLFF for volume-energy curve







Add to Input

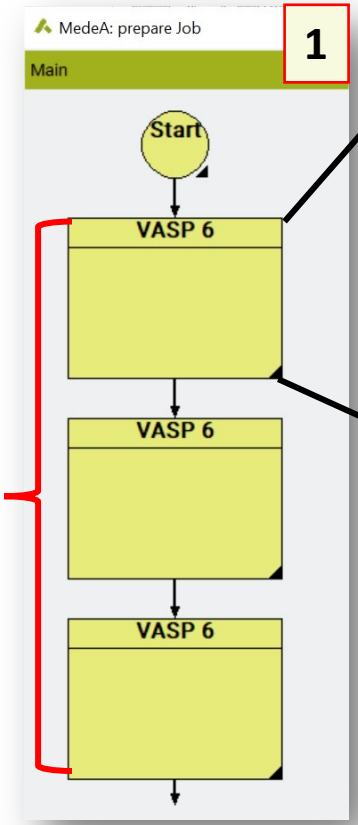


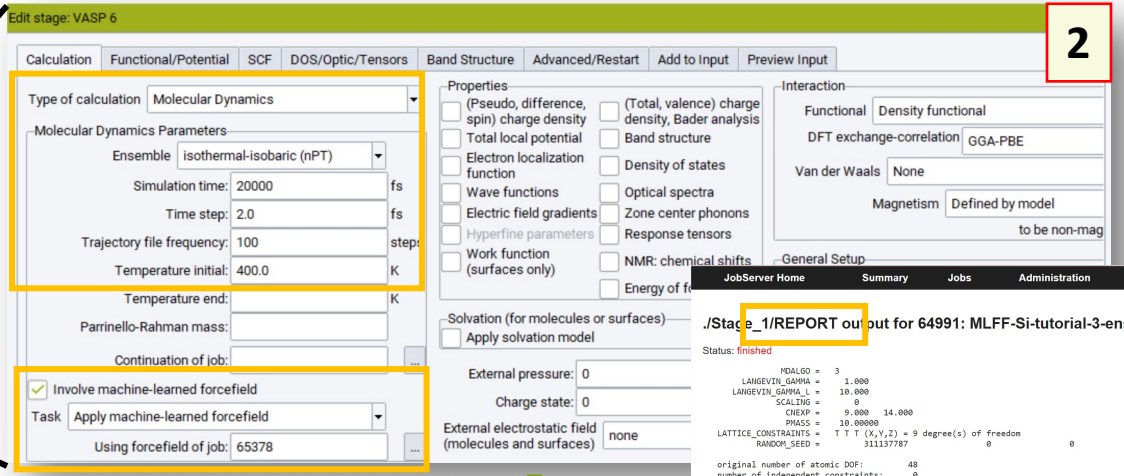
materials design

Apply trained MLFF for thermal expansion

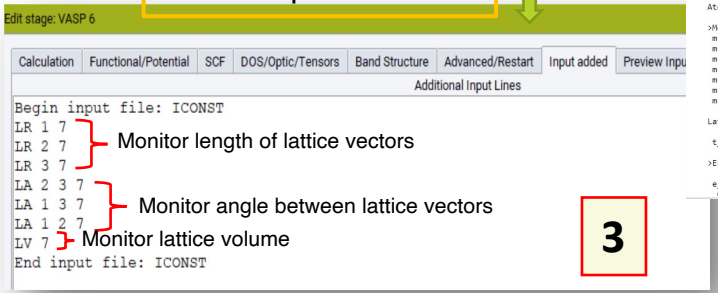
3 different calculations for 200 K, 300 K, 400 K

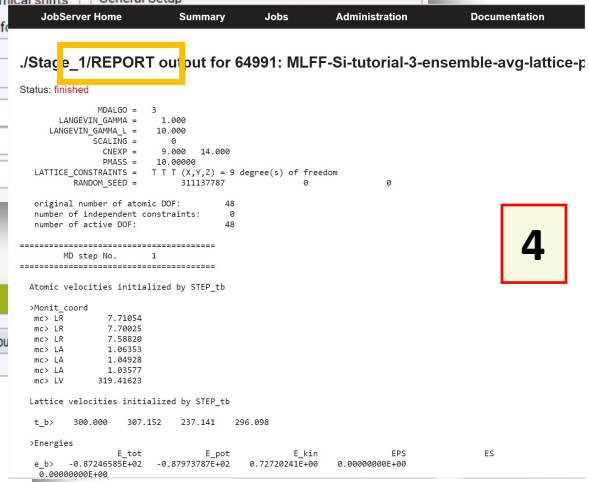
Ensemble average



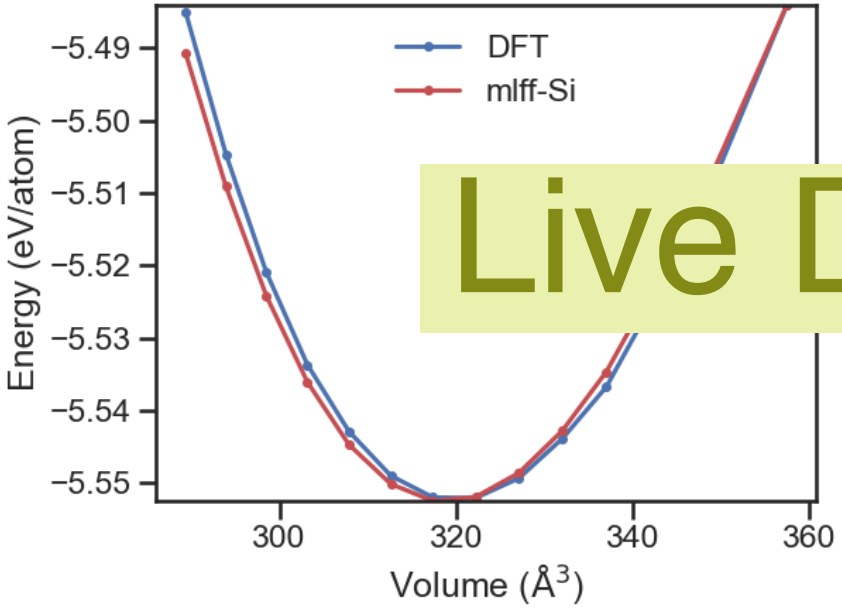


Add to Input: ICONST

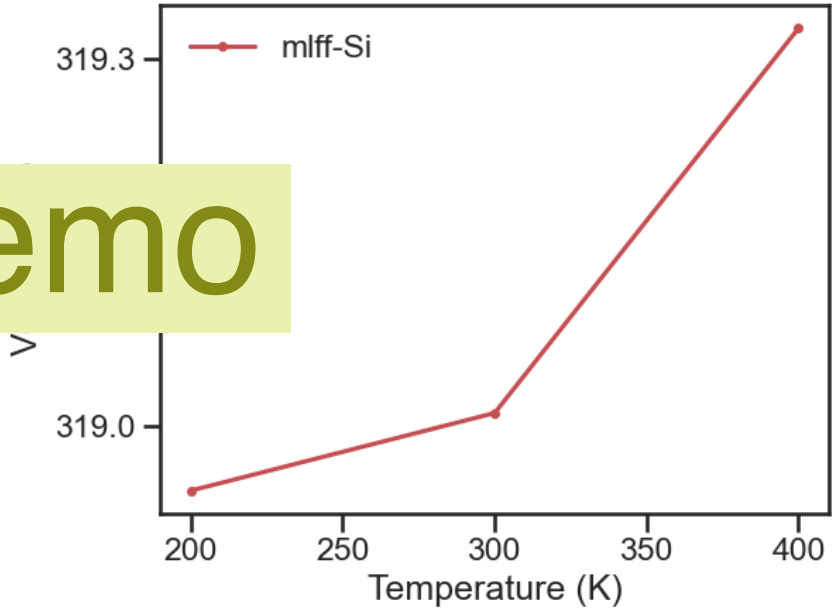




materials design



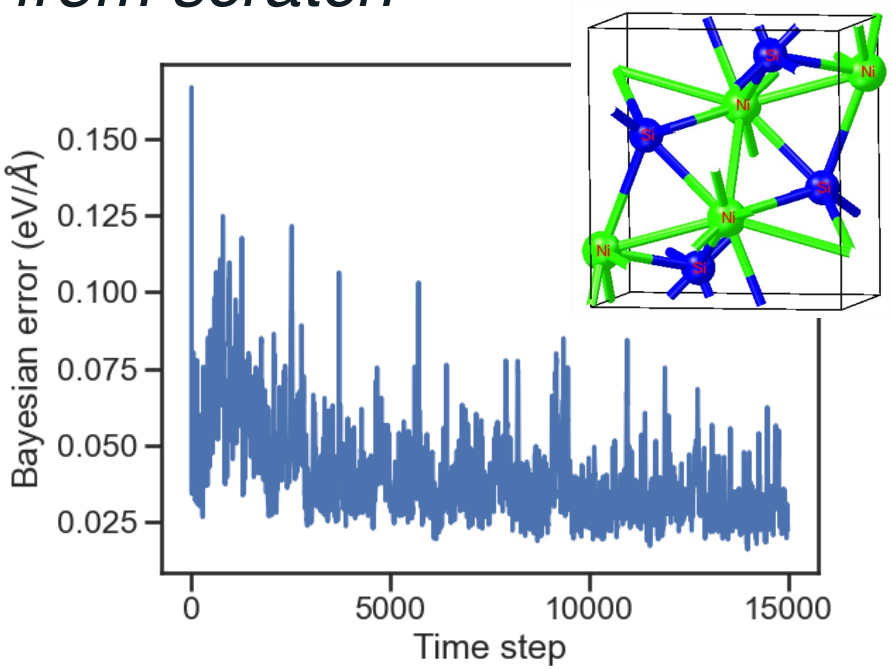
Live Demo

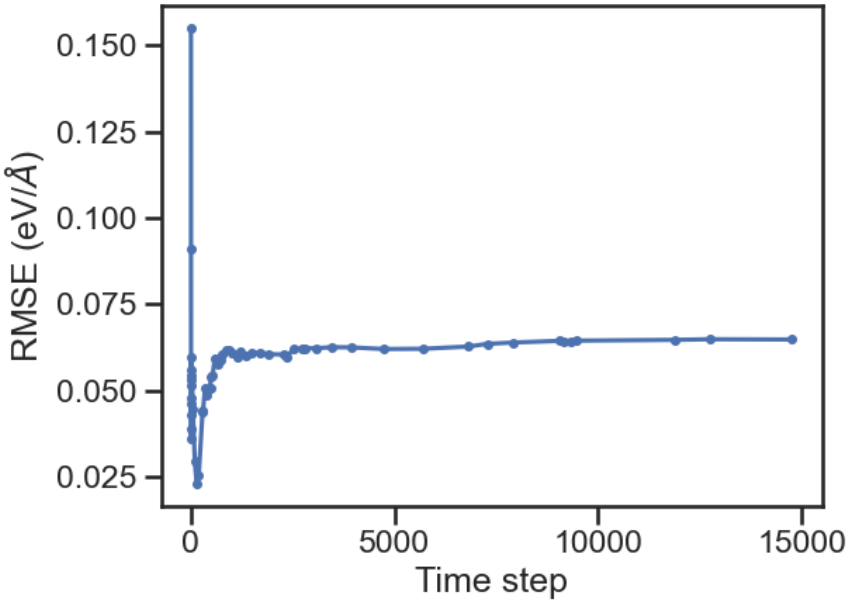


materials design

MLFF for NiSi: from scratch and starting from pretrained Si MLFF

(2) Pure NiSi: train an MLFF on NiSi *from scratch*



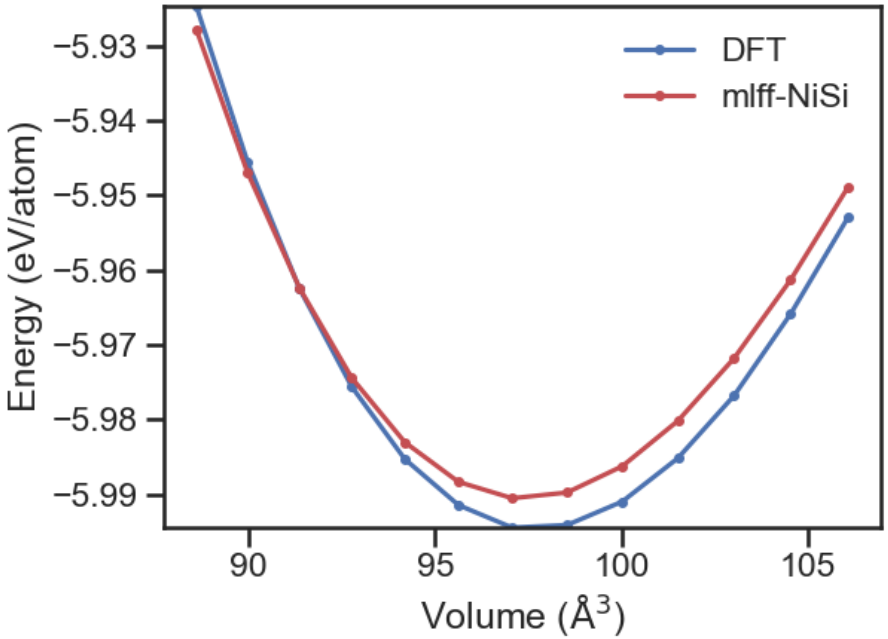


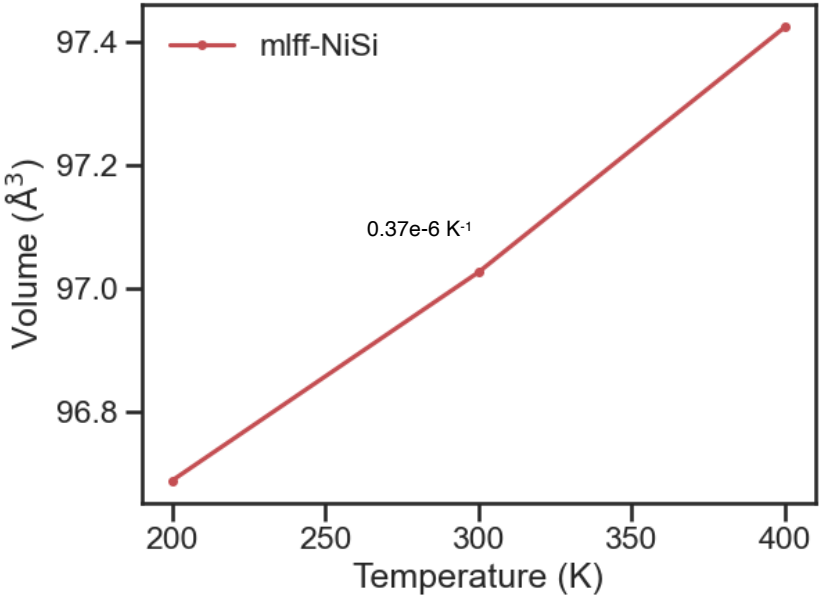
- 30ps, 400 K, nPT, unit cell used
- Acceptable accuracy on forces (<100 meV/Å)
- Improve fluctuations: more accurate DFT settings, supercell instead of unit cell
- Training time: ~45min, 8 atoms, 30 ps, 12 cores

materials design

(2) Pure NiSi – volume-energy curve and thermal expansion

MLFF trained from scratch on NiSi

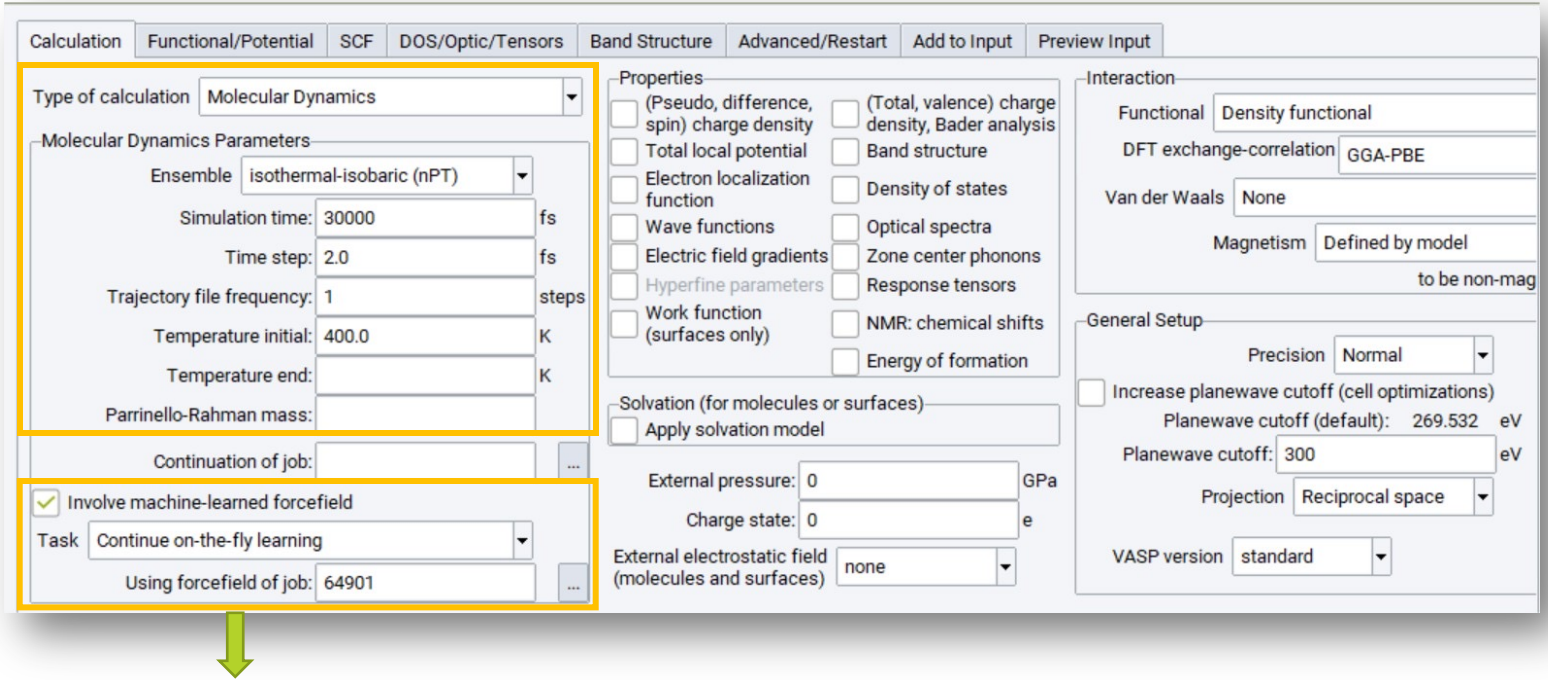




How to continue training from the pretrained MLFF on pure Si?

materials design

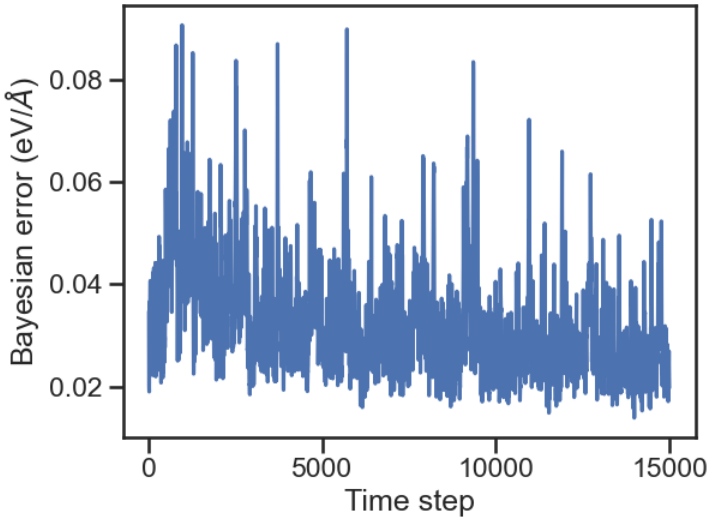
(2) Pure NiSi: train an MLFF on NiSi
starting from pretrained MLFF for Si

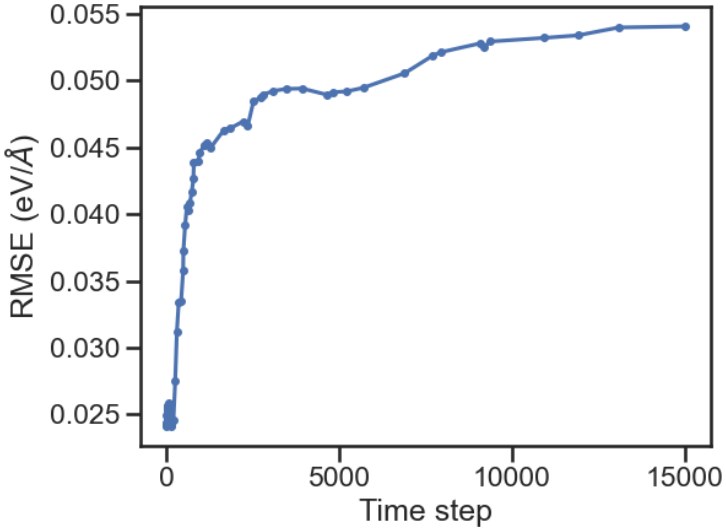


Continue from the MLFF trained on Si

materials design

(2) Pure NiSi - Bayesian error and RMSE on atomic forces *starting from MLFF of pure Si*



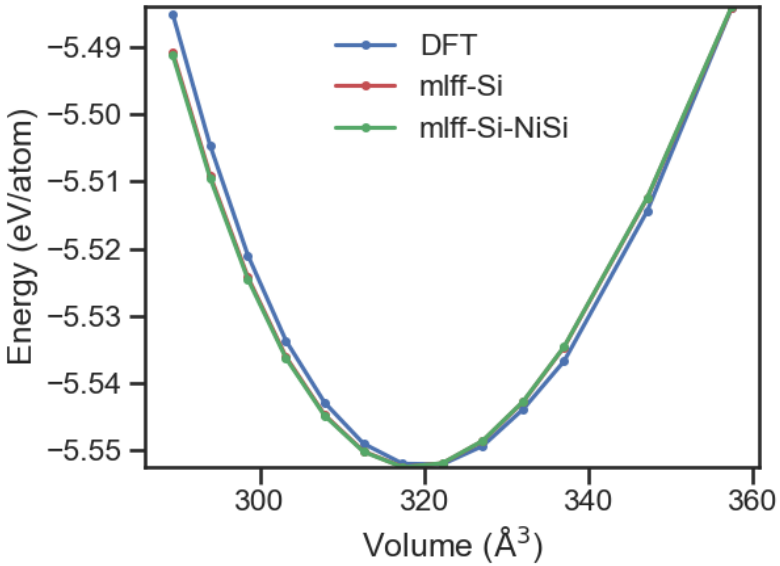


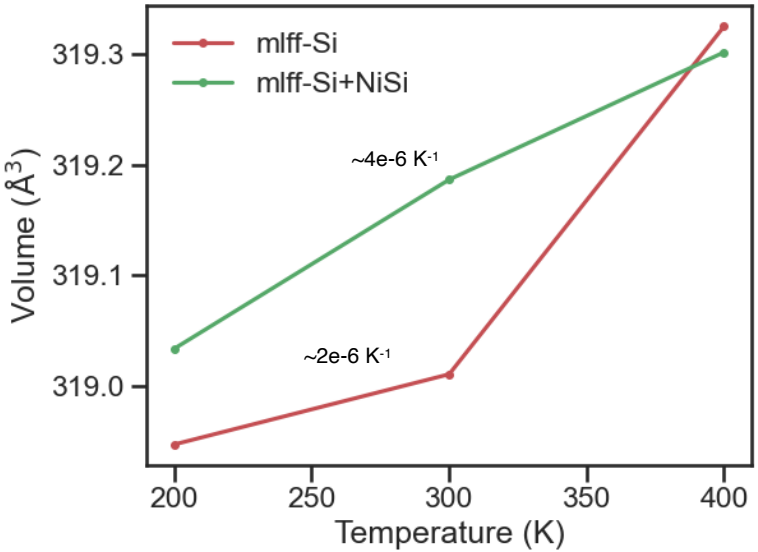
- 30ps, 400 K, nPT, 300 eV plane-wave cutoff, 2x2x2 k-mesh, unit cell used
- Larger fluctuations than MLFF trained on NiSi from scratch, but acceptable accuracy on forces (<100 meV/Å)
- Possible solutions to minimize fluctuations and error: higher plane-wave cutoff, denser k-mesh, larger cell size
- Training time: ~45min, 8 atoms, 30 ps, 12 cores

How does this combined Si+NiSi MLFF perform for Si and NiSi properties?

materials design

(1) Pure Si – volume-energy curve and thermal expansion *combined MLFF on Si+NiSi*

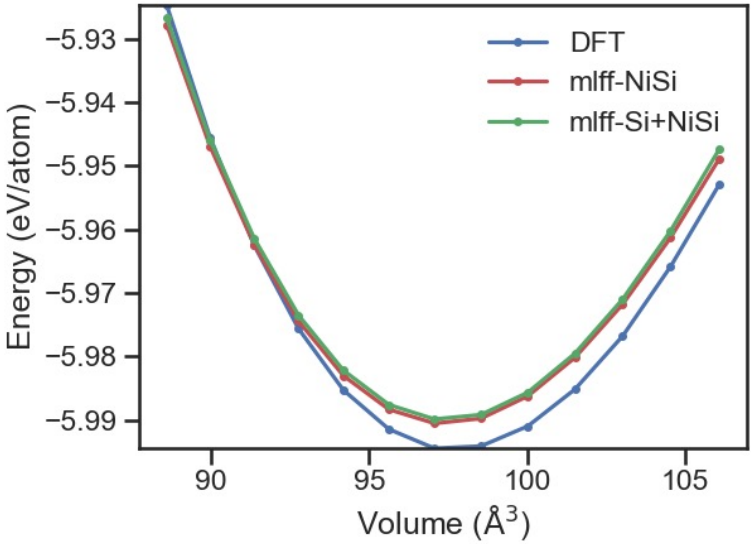


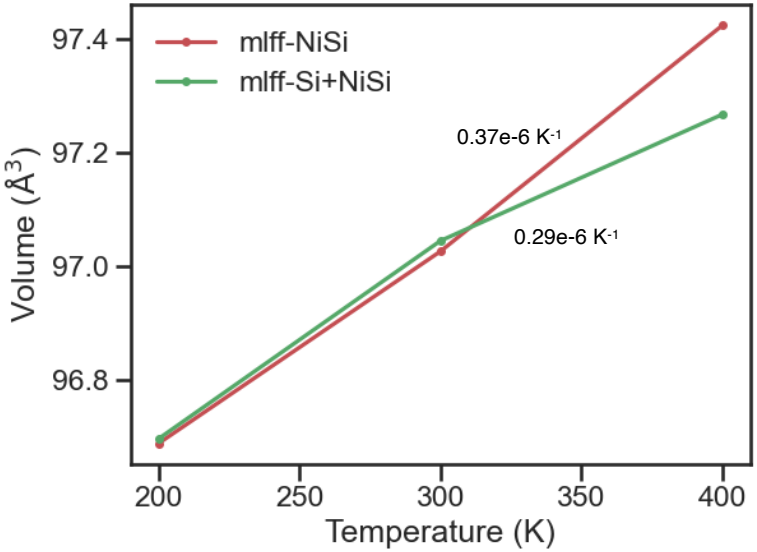


- MLFF trained on Si+NiSi performs equally good for volume-energy of Si, as MLFF trained on Si only
- Slight deviations between MLFF trained on Si+NiSi and Si only for thermal expansion of Si
- Compute time: ~11min, 16 atoms, 20 ps, 12 cores

materials design

(2) Pure NiSi – volume-energy curve and thermal expansion *combined MLFF on Si+NiSi*





- MLFF trained on Si+NiSi performs equally good for volume-energy of NiSi, as MLFF trained on Si only
- Slight deviations between MLFF trained on Si+NiSi and Si only for thermal expansion of NiSi at higher temperature
- Compute time: ~18min, 8 atoms, 20 ps, 12 cores

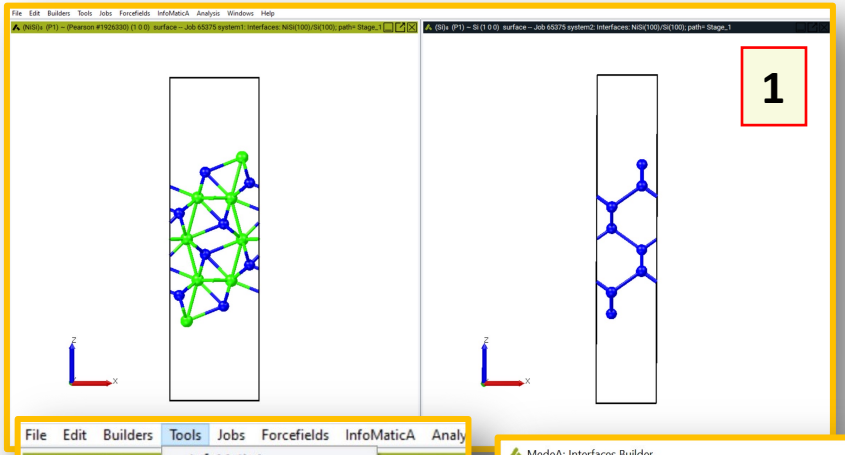
Next, an MLFF on Si/NiSi interface can be trained starting from Si+NiSi MLFF

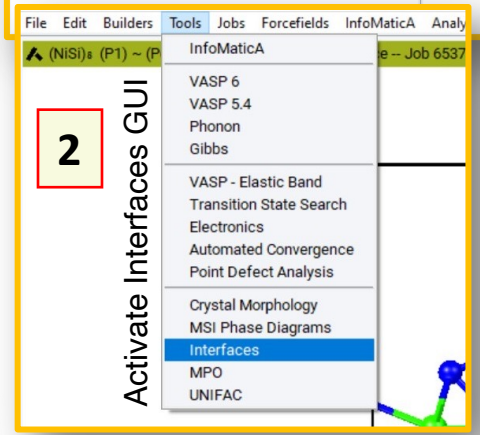
materials design

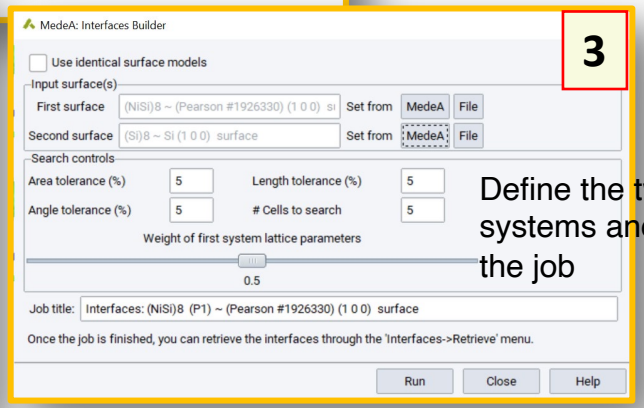
MLFF for Si/NiSi interface: starting from Si+NiSi MLFF

Si/NiSi (100) interface with *MedeA Interface Builder*

Build surface models

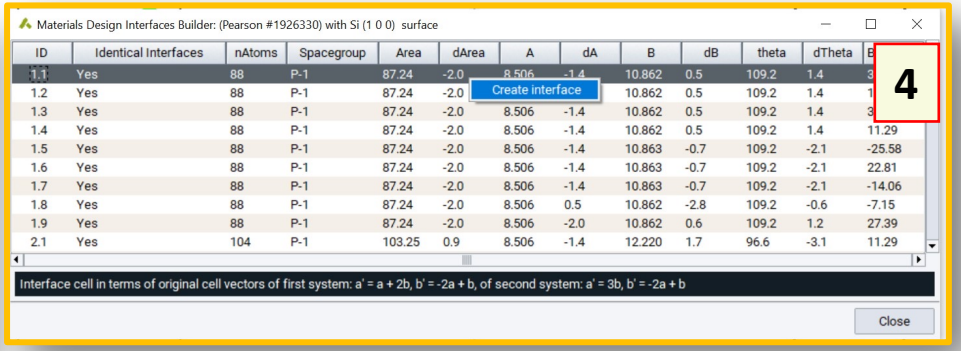


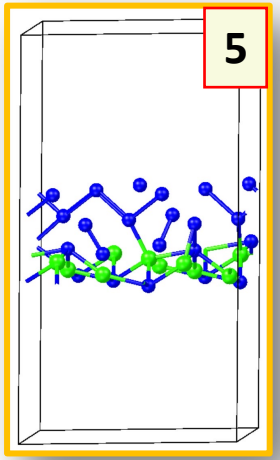




Define the two systems and run the job

Retrieve and create an Interface model

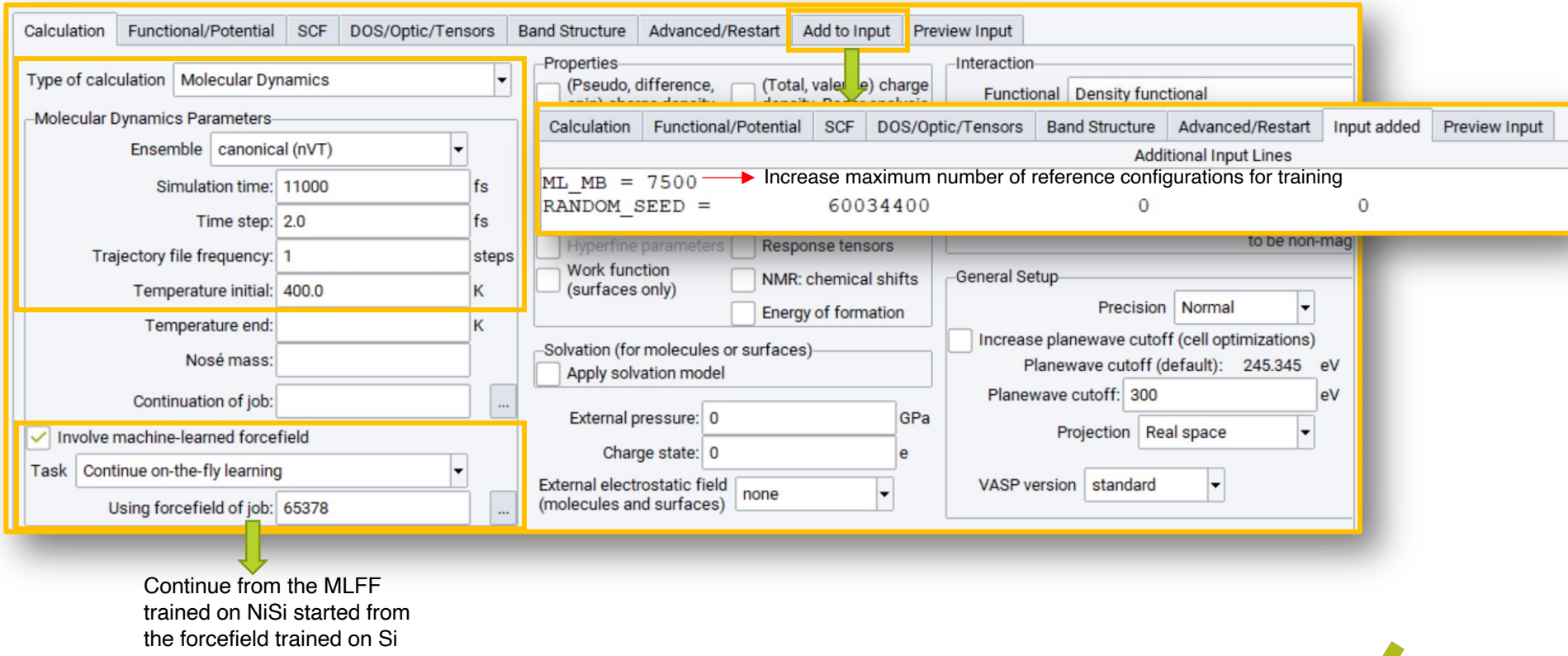




Si/NiSi (100) interface

materials design

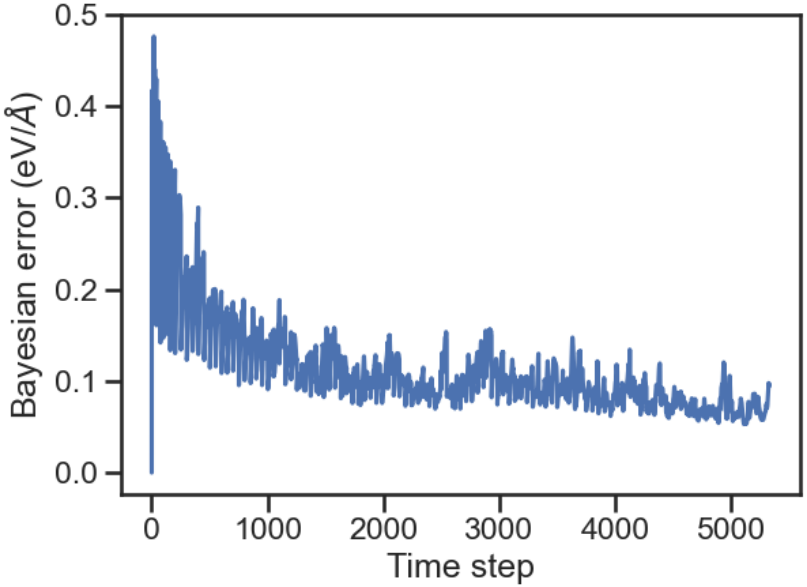
Train an MLFF on Si/NiSi interface
starting from MLFF trained on Si+NiSi

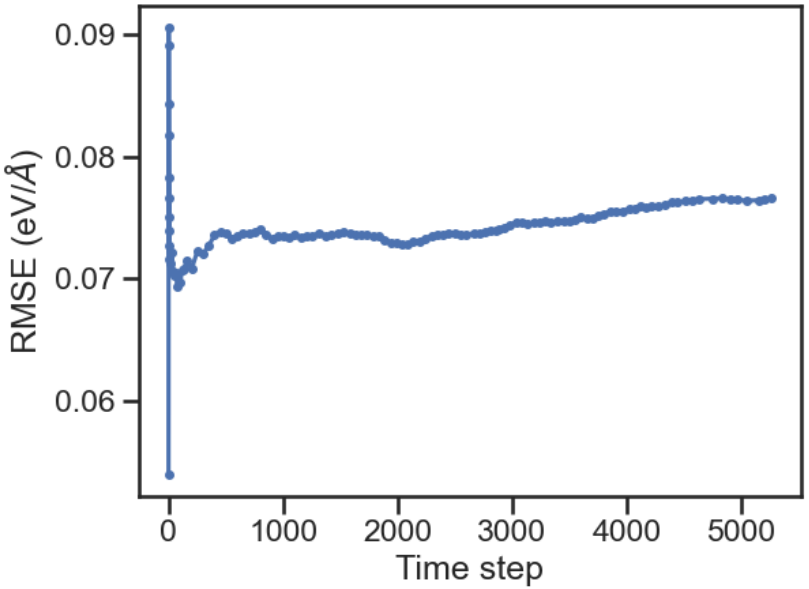


Continue from the MLFF trained on NiSi started from the forcefield trained on Si

materials design

Si/NiSi(100) interface- Bayesian error and RMSE on atomic forces *starting from MLFF of Si+NiSi*

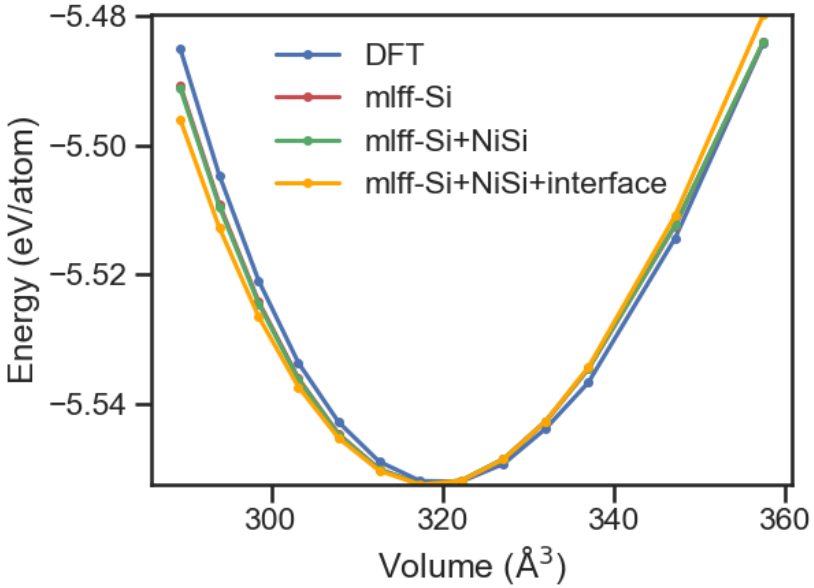


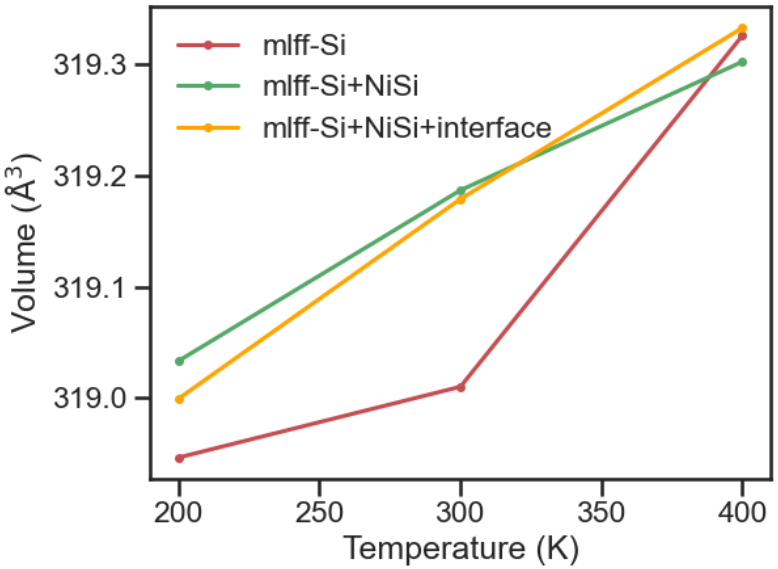


- 11ps, nPT, fixed cell shape (atom relaxation only), 300 eV cutoff, gamma-only point, 32-atom interface model
- Errors on forces slightly higher than previous MLFFs, due to complexity of the interface
- Possible ways to achieve better accuracy: train for longer than 10ps, more than 1 k-point in x-, and y-directions, use a thicker interface model
- Training time: ~9.5 hrs, 32 atoms, 11 ps, 12 cores

How does this combined Si+NiSi+interface MLFF perform for Si and NiSi properties?

Si: Si+NiSi+interface MLFF predicted volume-energy curve and thermal expansion

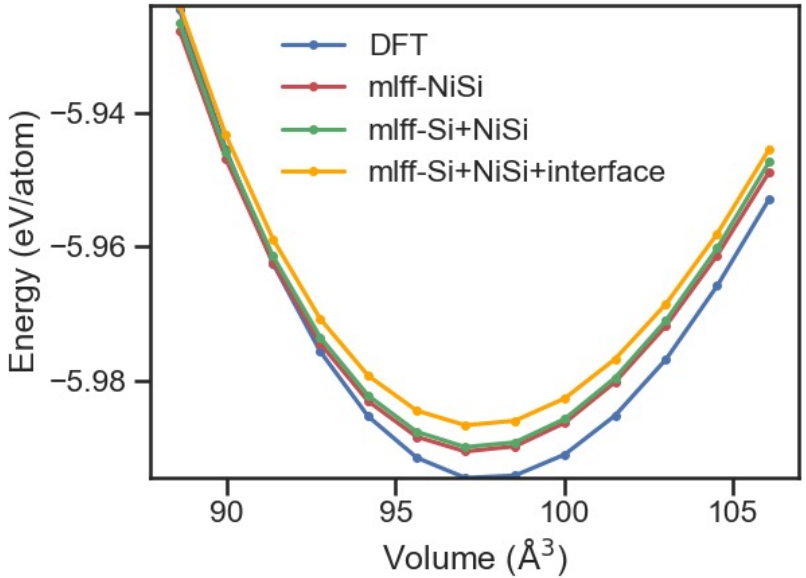


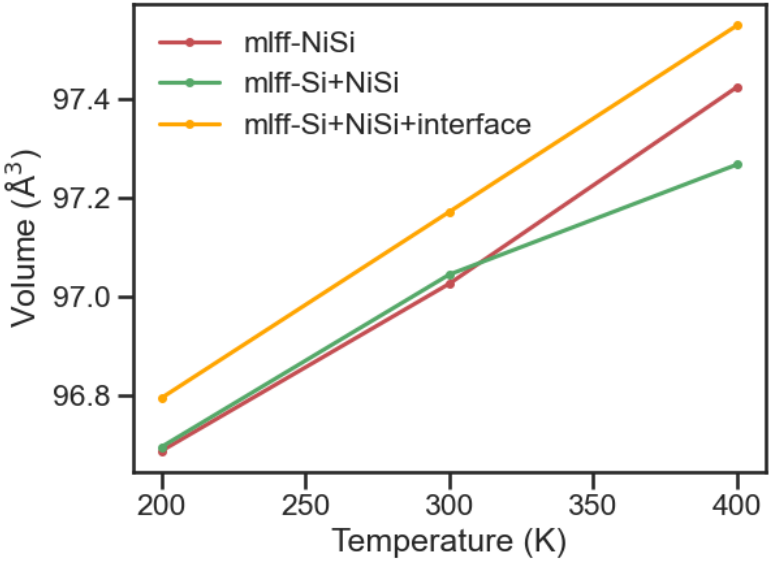


The combined Si+NiSi+interface retains the good agreement for volume-energy curves and behave similar to the Si+NiSi MLFF for thermal expansion

materials design

NiSi: Si+NiSi+interface MLFF predicted volume-energy curve and thermal expansion





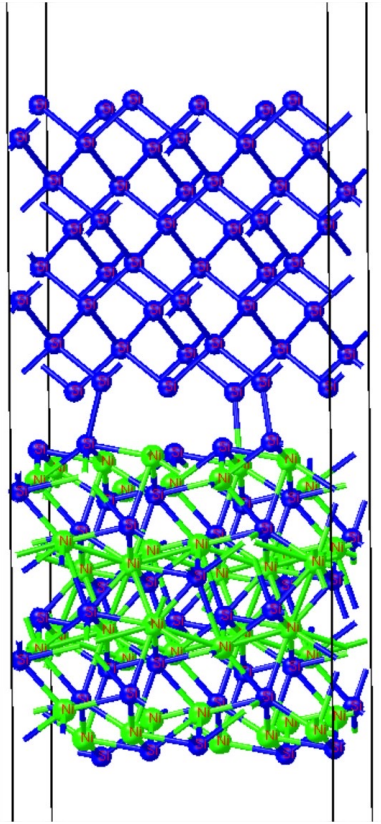
- Slight over-estimation of energies w.r.t DFT and other combined MLFFs, but all contained within 10 meV/atom
- Similar behavior for thermal expansion for all 3 MLFFs

materials design

Long MD simulation on Si/NiSi interface with combined MLFF

Apply the Si+NiSi+interface MLFF on a larger interface model

Si/NiSi (100) interface, 128 atoms



nVT ensemble, 20ps, 400 K

Calculation | Functional/Potential | SCF | DOS/Optic/Tensors | Band Structure | Advanced/Restart | Add to Input | Preview Input

Type of calculation: Molecular Dynamics

Molecular Dynamics Parameters

Ensemble: canonical (nVT)

Simulation time: 20000 fs

Time step: 2.0 fs

Trajectory file frequency: 100 steps

Temperature initial: 400.0 K

Temperature end: K

Nosé mass:

Continuation of job:

Involve machine-learned forcefield

Task: Apply machine-learned forcefield

Using forcefield of job: 68114

Properties

- (Pseudo, difference, spin) charge density
- (Total, valence) charge density, Bader analysis
- Total local potential
- Band structure
- Electron localization function
- Density of states
- Wave functions
- Optical spectra
- Electric field gradients
- Zone center phonons
- Hyperfine parameters
- Response tensors
- Work function (surfaces only)
- NMR: chemical shifts
- Energy of formation

Solvation (for molecules or surfaces)

Apply solvation model

External pressure: 0 GPa

Charge state: 0 e

External electrostatic field (molecules and surfaces): none

Interaction

Functional: Density functional

DFT exchange-correlation: GGA-PBE

Van der Waals: None

Magnetism: Defined by model

to be non-mag

General Setup

Precision: Normal

Increase planewave cutoff (cell optimizations)

Planewave cutoff (default): 269.532 eV

Planewave cutoff: 300 eV

Projection: Real space

VASP version: standard

Apply MLFF trained on Si/NiSi interface started from the forcefield trained on Si+NiSi

Compute time: ~13 hrs, 128 atoms, 20 ps, 12 cores

materials design

Guidelines for MLFF training

- Use a large enough system to avoid volume fluctuations and to “fit” collective vibrations into the supercell, to be able to apply MLFF to a larger supercell
- Important to learn correct forces: accurate DFT settings, dense k-points, 30% higher plane-wave cutoff than default, appropriate electronic convergence algorithm for good electronic convergence
- Do not use too large time-steps, especially for light elements
- When using temperature ramping, be careful to stay in temperature regime where no phase transitions are induced
- Use nPT ensemble whenever possible – additional cell fluctuations increases robustness of the resulting forcefield (perform nVT and refine MLFF)
- Test nPT settings: atomic and lattice friction coefficients with **LANGEVIN_GAMMA** and **LANGEVIN_GAMMA_L** tags (“Add to Input” in MedeA VASP GUI), and Parrinello-Rahman mass
- Increase Bayesian threshold for force if needed: **ML_CTIFOR** (“Add to Input”); default is 0.002 eV/Å

MedeA VASP MLFF and *MedeA MLPG*

- Both functionalities are machine-learned parameterizations of potential energy surfaces and are intended to calculate materials properties quickly with an accuracy of VASP
- In fact, VASP MLFF and SNAP MLP are both based on the Kernel Regression method
- However, MLPs (*SNAP and NNP*) are designed to work with much larger systems and with *LAMMPS*, while MLFFs are both trained and applied within VASP
- *MedeA MLPG* can use both VASP-MD trajectories and 0K calculated structures (distortions, defects, etc.), while VASP MLFF makes use of only MD trajectories for training

Relevant *MedeA* Modules

MedeA Environment (*InfoMaticA and Advanced Builders*)

MedeA VASP 6 (MLFF and DFT methods)

MedeA Interface Builder

MedeA HT-Launchpad (save an MD trajectory into structure list)

https://www.materialsdesign.com/webinars/recorded

materials design

Question and Answer Session

Dr. David Reith

Materials Design

Dr. Shubham Pandey

Materials Design

Dr. Xiaoli Liu

Materials Design

Announcements

ugm.materialsdesign.com

Next Week's MedeA Training

Thursday, October 27th

Dr. Marianna Yiannourakou

Materials Design

Dr. Jörg-Rütiger Hill

Materials Design

Next Week's Plenary Session

Tuesday, October 25th

Dr. Jozef Bicerano

Bicerano & Associates Consulting, LLC

Questions about Materials Design UGM

ugm@materialsdesign.com

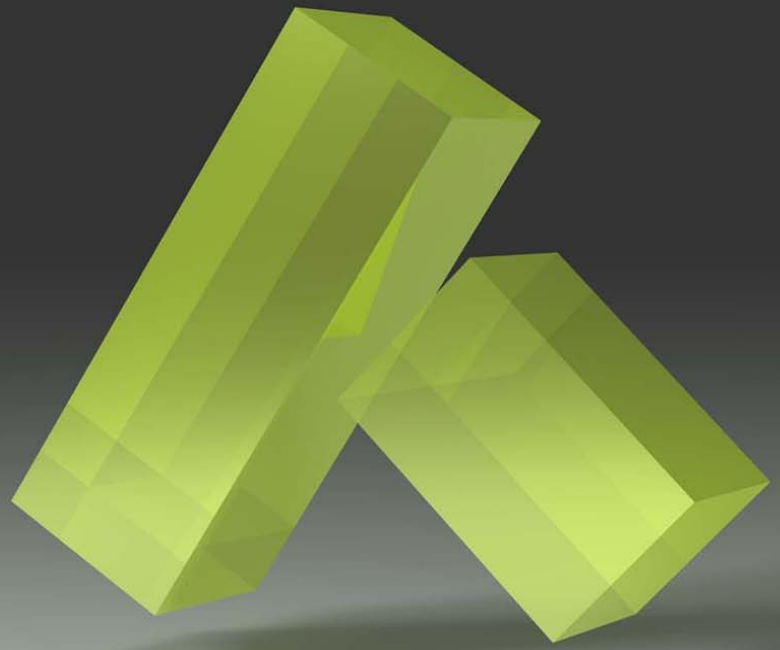
Support Team

support@materialsdesign.com

materials design

info@materialsdesign.com

www.materialsdesign.com



MedeA

Innovation by Simulation